

An Application of Least Squares Method to Analysis of Size Compositions

Yoshio FUKUDA

(Inland Sea Regional Fisheries Research Laboratory, Hiroshima, Japan.)

日本水産学会誌

第十九卷 第四号 別刷

昭和二十八年八月二十五日発行

Reprinted from Bulletin of
the Japanese Society of Scientific Fisheries

Vol. 19, No. 8, August, 1953

An Application of Least Squares Method to Analysis of Size Compositions

Yoshio FUKUDA

(Inland Sea Regional Fisheries Research Laboratory, Hiroshima, Japan)

Summary

This article presents a method of estimating mortality and growth parameters of a biological population through observed changes of the size composition, whether truncated or not. The principle is the least squares adjustment of data with estimation of parameters involved in the condition equations. A little complication comes out from the facts that the sample composition must be regarded as a vector variate, and that only some of the components are constrained to each condition equation. Though a little laborious in numerical calculation, it will especially meet such a situation where recruitment by growth is occurring and technically difficult for direct estimation. The number and combinations of the independently inferable parameters are considered, with which applicability is indicated to be generalized by adopting an appropriate division of size or time interval in advance. And a numerical example on a benthic larval population was discussed briefly for explanation and some interesting scopes of extrapolation are suggested.

Introduction

Almost completely covered by ordinary sampling schemes, the benthic population along the shore, less mobile or attached, does not seem to be so difficult as fish or plankton to estimate the standing crop and its change. However, as is the case with adults, we are obliged to face at a different situation when dealing with benthic larvae, for they pass their pelagic lives before setting on the ground and usually we cannot neglect possible recruitment from the pelagic population. In such cases we could start with estimation of the stock, but there seem to remain, other than sampling, some difficulties to be solved, that is, identification, counting, modes of precipitation and so on. Such a situation have we met in the course of study on seed collection of useful bivalves, where it is evident as well that we could not resort to the mere experiments without any population study in the fields.

This is a particular case to which we have intended to apply an arithmetic method of analysis presented below. It is always applicable to those cases where we could assure ourselves just in coverage of field surveys on successive occasions as well as representativeness of observed size-compositions, whether truncated or not. And, if desired, differential estimation of mortality and growth parameters by size

or/and period will be successfully designed. Somewhat laborious in computation, and consequently troublesome in appearance, as it is, it will often balance out technical complications otherwise required.

Outline of the Method

There are no particulars but some additional labors in calculation. In short, it is to adjust by least squares principle successive sample size compositions under some conditions, to which the parent ones should be constrained, and at the same time to estimate mortality and growth parameters involved in the condition equations [1].

1) **Estimation of Size Compositions** It goes without saying that nature of the problem makes it inevitable to adopt random sampling on successive occasions with sampling fraction as small as possible. Then every sampling unit quadrate for our case.....will include individuals to be possibly grouped into different size classes, that is to say, errors of class estimates within each sample composition would be no more independent each other. Consequently their covariances could not be neglected, while we might assume to be zero covariances between sample composition on different occasions. Additional labors in calculation come out really from these circumstances.

Now it is convenient to deal with a sample size composition on the occasion t as an k -dimensional vector variate, denoted by X_t , where k is the number of size classes. And the problem is to adjust sample compositions so as to minimize [2]

$$S = \sum_t (\bar{X}_t - \bar{m}_t) W_t (X_t - m_t),$$

where m_t is the parent composition vector corresponding to X_t , W_t is the weight matrix proportional to the inverse of the parent dispersion matrix, and throughout this paper the transposed of a matrix will be denoted by a bar on the letter.

2) **Condition Equations** Let us consider a simplified case, where a biological population is decreasing with a constant rate of mortality A and growing with a constant rate of growth B . Then the time-variation of the $(i+1)$ th class value, $\Delta_i m_{t, i+1}$ is composed of three parts; $A m_{t, i+1}$ will die and $B m_{t, i+1}$ grows up into the $(i+2)$ th size class, and at the same time $B m_{t, i}$ is recruited by growth from the i -th class. That is,

$$\Delta_i m_{t, i+1} = -(A+B)m_{t, i+1} + B m_{t, i}$$

After integration from t to $t+1$, replacing integrals by the trapezoid formula,

$$(1) \quad F^{i+1} \equiv m_{t+1, i+1} - m_{t, i+1} + (A+B)d_t(m_{t+1, i+1} + m_{t, i+1}) - B d_t(m_{t+1, i} + m_{t, i}) \equiv 0,$$

where d_t is half the time interval from t to $t+1$. This is one of the condition equation, to which the parent class values, $m_{t, i}$, $m_{t, i+1}$, $m_{t+1, i}$, $m_{t+1, i+1}$ must be constrained, so that the $(s-1)(k-1)$ condition equations will hold, when the size com-

positions, are obtained on s successive occasions. For linearization, neglecting terms of higher orders in the Taylor's expansion, we get the so-called reduced conditions as follows;

$$(2) \quad \mathbf{F}'_{m_t} \mathbf{x}_t + \mathbf{F}'_{m_{t+1}} \mathbf{x}_{t+1} + \mathbf{F}'_A \mathbf{e}_A + \mathbf{F}'_B \mathbf{e}_B = \mathbf{F}'_0(\mathbf{X}_t, \mathbf{X}_{t+1}, \mathbf{a}, \mathbf{b}),$$

where $\mathbf{m}_t = \mathbf{X}_t - \mathbf{x}_t$, $A = a - e_A$, $B = b - e_B$, and \mathbf{F}'_{m_t} , $\mathbf{F}'_{m_{t+1}}$, \mathbf{F}'_A and \mathbf{F}'_B are matrices made up of partial derivatives of \mathbf{F}' with respect to \mathbf{m}_t , \mathbf{m}_{t+1} , A , and B , that is,

$$\mathbf{F}'_{m_t}(k-1, k) \equiv \begin{pmatrix} -B & -1+A+B & 0 & \dots & 0 & 0 \\ 0 & -B & -1+A+B & \dots & 0 & 0 \\ 0 & 0 & -B & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1+A+B & 0 \\ 0 & 0 & 0 & \dots & -B & (-1+A+B) \end{pmatrix},$$

$$\mathbf{F}'_{m_{t+1}}(k-1, k) \equiv \begin{pmatrix} -B & 1+A+B & 0 & \dots & 0 & 0 \\ 0 & -B & 1+A+B & \dots & 0 & 0 \\ 0 & 0 & -B & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1+A+B & 0 \\ 0 & 0 & 0 & \dots & -B & (1+A+B) \end{pmatrix},$$

$$\mathbf{F}'_A(k, 1) \equiv \begin{pmatrix} m_{t,2} + m_{t+1,3} \\ m_{t,3} + m_{t+1,3} \\ \dots \\ \dots \\ m_{t,k} + m_{t+1,l} \end{pmatrix} \quad \mathbf{F}'_B(k, 1) \equiv \begin{pmatrix} m_{t,2} + m_{t+1,2} - m_{t,1} - m_{t+1,1} \\ m_{t,2} + m_{t+1,1} - m_{t,2} - m_{t+1,2} \\ \dots \\ \dots \\ (m_{t,k} + m_{t+1,l} - m_{t,k-1} - m_{t+1,l-1}) \end{pmatrix},$$

where the time factor is dropped, which will result in no confusion, though the parameters turn out to have a little different meanings, and read for the bracketed terms respectively $-1+A$, $1+A$, and $-m_{t,k-1} - m_{t+1,l-1}$ when all the larger than the k -th size class are pooled there.

3) Arithmetic For convenience of numerical calculation, procedures will be sketched here. Simply, assume sample size compositions on two successive occasions ($t=1,2$) to be at hand. Then simultaneous equations to be solved are, together with the condition equations (2) above introduced,

$$(3) \quad \mathbf{W}_1 \mathbf{x}_1 - \bar{\mathbf{F}}_{m_1} \mathbf{h} = 0 \quad \mathbf{W}_2 \mathbf{x}_2 - \bar{\mathbf{F}}_{m_2} \mathbf{h} = 0,$$

$$(4) \quad \mathbf{F}_A \mathbf{h} = 0 \quad \mathbf{F}_B \mathbf{h} = 0,$$

where \mathbf{h} is a $(k-1)$ dimensional column vector of Lagrange's multipliers.

Solving (3) by \mathbf{x}_t and inserting them into (2), we get

$$(5) \quad \mathbf{Lh} + \mathbf{F}_A \mathbf{e}_A + \mathbf{F}_B \mathbf{e}_B = \mathbf{F}_0,$$

where $\mathbf{L} = \mathbf{F}_{m_1} \mathbf{W}_1^{-1} \bar{\mathbf{F}}_{m_1} + \mathbf{F}_{m_2} \mathbf{W}_2^{-1} \bar{\mathbf{F}}_{m_2}$, a symmetric matrix and it is evident that these equations, together with equations (4), form the general normal equations. Then

the normal equations for parameters are

$$(6) \quad \begin{aligned} \bar{\mathbf{F}}_A \mathbf{L}^{-1} \mathbf{F}_A \mathbf{e}_A + \bar{\mathbf{F}}_A \mathbf{L}^{-1} \mathbf{F}_B \mathbf{e}_B &= \bar{\mathbf{F}}_A \mathbf{L}^{-1} \mathbf{E}_0 \\ \bar{\mathbf{F}}_B \mathbf{L}^{-1} \mathbf{F}_A \mathbf{e}_A + \bar{\mathbf{F}}_B \mathbf{L}^{-1} \mathbf{F}_B \mathbf{e}_B &= \bar{\mathbf{F}}_B \mathbf{L}^{-1} \mathbf{F}_0. \end{aligned}$$

Solving (6) we will get \mathbf{e}_A , \mathbf{e}_B , and then \mathbf{h} , \mathbf{x}_1 , \mathbf{x}_2 in the following way,

$$(7) \quad \begin{aligned} \mathbf{h} &= \mathbf{L}^{-1} (\mathbf{F}_0 - \bar{\mathbf{F}}_A \mathbf{e}_A - \bar{\mathbf{F}}_B \mathbf{e}_B) \\ \mathbf{x}_1 &= \mathbf{W}_1^{-1} \bar{\mathbf{F}}_{m_1} \mathbf{h} \\ \mathbf{x}_2 &= \mathbf{W}_2^{-1} \bar{\mathbf{F}}_{m_2} \mathbf{h}. \end{aligned}$$

Moreover, the minimized \mathbf{S} is evaluated as follows;

$$(8) \quad \begin{aligned} \text{Min } \mathbf{S} &= \sum \bar{\mathbf{x}} \mathbf{W}_i \mathbf{x}_i \\ &= \bar{\mathbf{h}} \mathbf{F}_0 \\ &= \bar{\mathbf{F}}_0 \mathbf{L}^{-1} \mathbf{F}_0 - \bar{\mathbf{F}}_0 \mathbf{L}^{-1} \bar{\mathbf{F}}_A \mathbf{e}_A - \bar{\mathbf{F}}_0 \mathbf{L}^{-1} \bar{\mathbf{F}}_B \mathbf{e}_B. \end{aligned}$$

Accordingly, we can evaluate $\text{Min } \mathbf{S}$ even without obtaining \mathbf{h} , but it will be desirable for us to solve out to residuals of size compositions for further discussion.

Now we have the adjusted size compositions and the estimated rates of mortality and growth at hand,

$$\begin{aligned} \hat{\mathbf{m}}_t &= \mathbf{X}_t - \mathbf{x}_t \quad (t=1, 2) \\ \hat{A} &= a - e_A \\ \hat{B} &= b - e_B. \end{aligned}$$

The standard errors of these estimates will have been obtained in the course of calculation, as explained in the next paragraph.

4) **Nature of Solution** Approximately as usually adopted, we will start numerical computation by substituting sample and some guessed values into \mathbf{F}_A , \mathbf{F}_B , and \mathbf{F}_{m_t} , with the consequent problem of convergence untouched. Then the basic model in the statistical sense is specified by

$$\begin{aligned} \varepsilon(\mathbf{X}_t - \mathbf{m}_t) &= \varepsilon(\mathbf{E}_t) = 0 \\ \varepsilon(\mathbf{E}_t \bar{\mathbf{E}}_t) &= \mathbf{w}_t^{-1} \sigma^2 \quad (t=s) \\ &= 0 \quad (t \neq s), \end{aligned}$$

and for the present we can assume constant errors of the guessed values for parameters,

$$E_A = a - A, \text{ and } E_B = b - B.$$

Then the expectation of

$$\mathbf{F}_0 = \mathbf{F}_{m_1} \mathbf{E}_1 + \mathbf{F}_{m_2} \mathbf{E}_2 + \mathbf{F}_A E_A + \mathbf{F}_B E_B$$

is $\mathbf{F}_A E_A + \mathbf{F}_B E_B$ and the parent dispersion matrix of \mathbf{F}_0 becomes \mathbf{L} , where \mathbf{L} is the same as already obtained in the course of calculation.

Denoting the inverse of the coefficient matrix of equations (6) by $\mathbf{C}(c_{ij})$,

$$\begin{aligned} \mathbf{e}_A &= (c_{AA} \bar{\mathbf{F}}_A \mathbf{L}^{-1} + c_{AB} \bar{\mathbf{F}}_B \mathbf{L}^{-1}) \mathbf{F}_0 \\ &= (c_{AA} \bar{\mathbf{F}}_A \mathbf{L}^{-1} + c_{AB} \bar{\mathbf{F}}_B \mathbf{L}^{-1}) (\mathbf{F}_{m_1} \mathbf{E}_1 + \mathbf{F}_{m_2} \mathbf{E}_2) + E_A. \end{aligned}$$

Then the adjusted estimate \widehat{A} is expressed as:

$$\begin{aligned}\widehat{A} &= a - \mathbf{e}_A = A + E_A - \mathbf{e}_A \\ &= A - (C_{AA}\bar{\mathbf{F}}_A\mathbf{L}^{-1} + C_{AB}\bar{\mathbf{F}}_B\mathbf{L}^{-1})(\mathbf{F}_{m_1}\mathbf{E}_1 + \mathbf{F}_{m_2}\mathbf{E}_2),\end{aligned}$$

the expectation of which is therefore A and similarly $\epsilon(\widehat{B}) = B$. In this sense, that is, with possible problem of convergence set aside, the guesswork at the start has nothing to do with the unbiased character of the adjusted estimates. The dispersion matrix of them is given by \mathbf{C} , where \mathbf{C} , as denoted above, is already obtained in the course of solution.

In addition, the expectation of Min S turns out to be

$$\begin{aligned}\text{Min } S &= (2k-2)\sigma^2 \\ &= (sk-p)\sigma^2,\end{aligned}$$

in general, where p is the number of parameters estimated. It is easily understood that this relation will give us the so-called external estimate of σ^2 . Exact test of goodness of fit needs an explicit form of probability distribution to be introduced, while an approximate inference could be made by comparison of this estimate with the internal one.

How many parameters to be possibly estimated

So far, we have assumed a biological population with constant rates of mortality and growth. In practice, however, there will often arise such cases where parameters are suspected to change by size or/and by period. The procedure of solution is easily extended to those cases, but we cannot estimate all parameters to be possibly involved in the condition equations.

The number of parameters, independently inferable, is determined by the solvability condition of the general normal equations. The necessary and satisfactory condition for presence of a unique solution is that the determinant of their coefficients is not zero. From this condition it follows at once that we can independently estimate at most no more than $(s-1)(k-1)$ parameters, though possibly involved in the condition equations are totally $2(s-1)(k-1)$ when all the lagers are pooled in the largest size class, or $(s-1)(2k-1)$ when they are truncated, because the rank of the submatrix \mathbf{L} is at most $(s-1)(k-1)$. Moreover, it does not imply whatever combination of $(s-1)(k-1)$ parameters, if desired, is under our choice. For, in that case, where $(\mathbf{F}_A, \mathbf{F}_B)$ turns out to be a quadratic matrix, an additional condition must be satisfied that the determinant $|\mathbf{F}_A, \mathbf{F}_B| \neq 0$. In the same manner, when the number p of parameters chosen is less than $(s-1)(k-1)$, a condition must hold, that is, the rank of $(\mathbf{F}_A, \mathbf{F}_B)$ has to be equal to p . Reversely speaking, these conditions are nothing but that we must introduce some more assumptions about possibly involved parameters, that is, in other words, that our choice

of the biological structure to be fitted is restricted. It is true, but these circumstances imply the least limitation in the applicability of this method, because those difficulties would be successfully overcome in advance by adopting an appropriate division of size or time intervals at which the parameters to be assumed constant.

Some more definitely, let us begin with the two-successive-sample case. Remembering that we have in any case to determine one mortality and one growth parameters, it is evident that we require at least three size classes for application of this method. With one more size class added, the parameters possibly involved in the conditions increase by two, one mortality and one growth, while the independently inferable do just by one. To estimate three parameters, there must be four or more size classes present, when no cares are required for their combinations. In the case of four parameters, with five or more size divisions necessary, any other combinations will be under our choice but some of the ones of two mortality and two growth. The unfavorable ones involve, as different from the other correspondings, the growth parameter of the second, or, in the truncated case only, the mortality and growth parameters of the largest size class. Five parameters, independently inferable, which require six or more size classes present, will show another feature. Then we could choose any combination of one mortality and four growth, or vice versa. However, of others, that is, two mortality and three growth or vice versa, those partly including one or both of unfavorable combinations in the four-parameter case are impossible. Moreover, out of our choice are, among combinations of three mortality and two growth, those which involve the mortality parameters of the successive two size classes and the growth one of the smaller, as different from the others. In this way we could do enumeration of favorable or unfavorable combinations, following to the above conditions, without much difficulty, and case by case in practice, though the more the independently inferable, the more complicated it may be.

Next, a word to the three-successive-sample case. It is easily to be understood that we can estimate as different by period all parameters involved in the favorable combination of the two-successive-sample case. And in this case, it is evident as well, two size classes are sufficient to determine one mortality and one growth parameters. This fact implies that the unfavorable circumstances in the former case could be removed by assuming the parameters constant over both period. Otherwise, that is to say, when the parameter combination of at least either period is unfavorable, we cannot solve the problem.

Another point to be considered along this line is what formula we had better use in replacing integrals in the conditions, though we have adopted above the trapezoid one. Simpson's formula, for example, will surely improve condition equa-

tions in accuracy, while it is to be taken into account that we are obliged in return to presume parameters to be constant over the longer period. How much it pays us will depend, of course, upon situations and our experiences.

Numerical Example

This is a part of the field survey on *Venerupis* around the end of Hiroshima Bay at the mouth of the river Yahata, carried out by the Hiroshima Experimental Fisheries Station. The survey repeatedly covered the seven plots (each about 340 square meters in area) on five successive occasions during May to July 1952. Here will we refer only to the results of the plot A on two successive occasions, 6th and 24th June.

They used the 5×5 cm² quadrat (5 cm deep) as sampling unit and 34 units per plot, with sampling fraction of about 1/4,000, were drawn at random on each occasion. Total fraction through the period is still so small that we could neglect the finite correction for the variance estimates. Bottom samples were filtered by five different meshes, the sizes of which are 1,2,3,4, and 5 mm, and *Venerupis* larvae were counted size by size. It is evident that all the larger than 5 mm size class are naturally pooled there.

Equal sampling fraction allows us to adopt sample totals as the estimated size compositions. The sample dispersion matrix is given by $nD/(n-1)$, where n is the size of sample and D is composed of sample sums of squares and products (Table 1). Take the weight inverse to this matrix. Then we could do without inversion of it, because it is the inverse of the weight matrix that is necessary in computation. In addition, we must make some guess of parameters. We will take $a=0.10$ and $b=0.60$ in this example. The closer to the true values we could guess them, the more we would gain from a view-point of convergence. Otherwise they might result in the more cycles of operations to be continued for better estimation. Necessary matrices written out before, will be easily made up of these guessed values and sample compositions.

Table 1. Materials from the Survey

Class Orders	Sample Size Composition	Matrix* of Sample Sums of Squares and Products				
6th June						
1	534	6712.91	7741.27	1172.27	116.91	47.91
2	1291		17663.52	2735.18	250.94	82.94
3	609			3099.18	329.27	-12.73
4	116				142.24	-6.76
5	17					14.24

24th June						
1	17	22.94	85.64	225.18	43.73	5.12
2	486		3230.55	2396.73	459.91	94.18
3	1176			10081.64	983.55	349.09
4	282				772.18	47.36
5	91					122.05

* Symmetric to the diagonal. so the elements below are omitted.

Now, we are ready to run a computational machinery. Complicated as in appearance, it will be almost automatically operated, following to the procedures above outlined. After some multiplications, additions, and reversions repeated back and forth, with a cycle of operations completed, we have a solution at hand. The adjusted estimates of parameters, together with minimized S, are shown in the second column of Table 2.

Some Discussions

In this example, we have started out with $\sigma^2 = n/(n-1)$, as internal estimate, while the external estimate, to be noted, has come out to be about five times larger. Distribution-freely proceeding, we know no exact level by which to test significance of this ratio, but it appears to be a little high. In other words, goodness of fit is just to be doubted of the assumed biological model. Before examination of it, however, there is another possibility present to be considered. It is evidently related to the guessworks at the start, and, really for this example, they are largely adjusted, an indication that they might have been inadequate. To throw light on this point, the results of successive cycles of calculations are presented in Table 2. Their convergence will depend upon our guesswork, it is true, but the adjusted

Table 2. Results of Successive Cycles of Calculations

	Guesed Values	First Cycle	Second Cycle	Third Cycle	Fourth Cycle
Parameters	0.10	0.200	0.169	0.149	0.129
Mortality		0.057	0.056	0.056	0.057
S. E.		0.293	0.232	0.205	0.196
Growth	0.60	0.031	0.024	0.022	0.022
S. E.					
Minimized S	—	39.834	47.360	63.324	65.267

estimates seem, after a few cycles, to converge within orders of the estimated standard errors. Successfully their convergence might happen to result in reduction of the external estimate of σ^2 . Otherwise, as in Table 2, where significance turns out to be more assuring, we must take a step farther to inspect the biological model to be presumed. Then residuals of condition equations will give us some clues to find out where and how the alternative should be improved. In this example,

after the third cycle, they are in order as follows,

$$-288.2; 834.8; -59.5; 8.4.$$

The first two, related to the first three size classes, are conspicuous in discrepancy and, to be noted here, their absolute magnitudes vary inversely to each other, as parameters decrease. Therefore it may be better to assume different parameters between the first two and the other size classes.

Table 3 sums up the results of three trials, continued from the above third cycle, that is, with uniform mortality of 0.149 and growth of 0.205 at the start. There it is observed, first of all, that either alternatives are expected to improve

Table 3 Summary of Three Alternatives

Model 1	Two Mortality and One Growth Parameters	
Mortality for the smaller	0.395	Growth 0.076
" for the larger	-0.275	
Residuals of Conditions	-9.9; 76.7; -38.9; 14.1	
Model 2	One Mortality and Two Growth Parameters	
Mortality	0.058	Growth for the smaller 0.469
" for the larger	0.158	
Residuals of Conditions	-126.9; 119.1; -30.0; -56.6	
Model 3	Two Mortality and Two Growth Parameters	
Mortality	0.295	Growth 0.208 for the smaller
"	-0.188	" 0.096 for the larger
Residuals of Conditions	-25.8; 33.3; -42.0; 15.5	

goodness of fit. And at the same time it is to be noted that Model 1 and 3 will give us a negative rate of mortality for the larger size interval. Generally speaking, negative mortality indicates presence of recruitment of different feature from the one by growth. This is the case or not, with our example, is of course out of the statistical problem, and we must remember, reversely, too, that it is another problem that the statistically fitted model is biologically approved or not. With our case,

Table 4. Summary of Completed Analysis for Model 2, with One Mortality and Two Growth parameters

Size Composition		1	2	3	4	5
6 th June	Observed	534	1294	609	116	17
	Adjusted	533	1286	591	113	18
24 th June	Observed	17	486	1176	282	91
	Adjusted	18	505	1145	310	81

Parameters	Mortality throughout constant	0.064	$10^{-3} \times \begin{pmatrix} 3.3525 & 0.6937 & 0.7207 \\ & 1.6629 & 0.4270 \\ & & 0.4359 \end{pmatrix}$
	Growth for class 1 and 2	0.542	
	Growth for class 3, 4 and 5	0.165	
Minimized S		2.2055	

* Because of symmetry, the elements below the diagonal are omitted. And the internal estimate of σ^2 is replaced approximately by 1.

such a recruitment seems to be possibly improbable, though it is to be discussed in details from various points of view. And, for the present, we will adopt Model 2. Then, another cycle of operations leads us to the solution, shown in Table 4. It shows that the external estimate is significantly reduced. And it is easily understood that possible improvement is statistically tested by comparison of minimized *S*'s equivalently to the principle of analysis of variance. In this way it is undoubtedly evident that the observed change of the size composition, apparently simple as it is, is really not so straight-forward, and a critical change of parameters takes place around 3 mm.

Three Extrapolations

Extrapolations, always in danger of erroneous inference, may not be so desirable. However, we are here interested in the three aspects of them. First of all, it is evident that we could, by extending the model to the smaller size class, infer to deposition of larvae, or, so to say, an effective recruitment for the period from the pelagic stock. Backward extrapolation, and forward one as well, has a little difficulty, for condition equations are lacking by one for the purpose of estimating a previous, or a future, size composition. However, with rate of growth already estimated, we could measure approximately the time elapsed before individuals of a size class totally grew up to the present class, that is, before complete displacement of a size class. As a consequence, size compositions on particular occasions, thus measured, will be derived easily by multiplying appropriate survival factors to the present status. Different rate of growth will bring in some complications, for example, discrepancy of the particular period just cited. Then, however, our condition equations modified will help connecting those independent processes. Thus we could make an inference, roughly as it is, for neither individual differences of growth nor possible changes of parameters are taken into account, when deposition, or spawning, began, and how pelagic larvae ever since have been effectively recruited to the benthic population. As for forward inference, we must make an assumption about recruitment, for example, that it has just ceased. Then we could estimate a size composition on an arbitrary future occasion. Though the variances of these extrapolated estimates are approximately evaluated by the ordinary formula about the variance of a function of variates, we must remember that they are no safeguard against erroneous inference we might happen to fall in on extrapolation.

With detailed discussion left over to a future publication, here will be a few points referred to. Both parameters, originally defined as instantaneous rate, turn out to be first approximations to the ordinary rates integrated over half the period when the time factor is included into them. Taking care of this difference, it follows from

the results of Model 2 that the effective recruitment from the pelagic stock, small as it is at the early stage, seems to have begun five months before and then gradually increased, and that the truly effective one has taken place since about five or seven weeks before. Assuming, as expected from the observed change of the size composition, that recruitment has just ceased after the period in this example, let us estimate a size composition on the occasion when the same plot was subsequently covered, and then we have found a striking coincidence between the observed and the calculated size compositions as follows;

Date	Class Order	1	2	3	4	5
7th July	Observed	1	139	1020	600	194
	Calculated	0	202	1056	448	161

This coincidence, we could not deny, is an evidence for adequacy of Model 2.

A Note

This may be unnecessary, but will not be entirely irrelevant, for there is probably expected a situation, where we have no more than single size composition on each occasion, or no information available on the dispersion and accordingly on the weight, with which we would be hardly confronted in a designed survey. Strictly speaking, there seems to be no way to meet such a situation. What is usually adopted is to neglect covariances and to weight equally, or some improvedly, inversely to the class frequencies. Upon this substitute, on account of apparent gain of reducing computational labors, we are so often inclined to rest.....more or less idly. However, though it is true that it depends partly on experiences and future study, these substitutes will not be so desirable, because they are based upon complete disregard of operation particulars of the survey itself and that covariances could not be always neglected from a biological point of view, as really in this example.

Closing this article, the author wishes to express many thanks to Dr. T. Hanaoka, Director of the Laboratory, by whom he has been continuously encouraged to advance this study, and to Mr. T. Fujita, chief of the Hiroshima Experimental Fisheries Station, and his staffs, who have gladly provided him with materials of the survey for explanatory discussion in this article.

References

- [1] DEMING, W. E. (1946)
Statistical Adjustment of Data. JOHN WILEY and Sons, New York. Translated into Japanese by S. MORIGUCHI
森口繁一, 推計学によるデータのまとめ方 1951年 岩波書店
- [2] MASUYAMA, M. (1948)
Note on a Markoff's Theorem of Least Squares (in Japanese) Contr. Math. Stat. Tokyo, No. 11, Vol. 4. 統計数理研究所 講究録

* Chief of the Nagasaki Experimental Fisheries Station since December 1952.